

+

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): MICHAEL P. CLEMENTS and ANA BEATRIZ C. GALVÃO

Article Title: TESTING THE EXPECTATIONS THEORY OF THE TERM STRUCTURE OF INTEREST RATES IN THRESHOLD MODELS

Year of publication: 2003

Link to published

version: <http://dx.doi.org/10.1017/S1365100502020163>

Publisher statement: None

TESTING THE EXPECTATIONS THEORY OF THE TERM STRUCTURE OF INTEREST RATES IN THRESHOLD MODELS

MICHAEL P. CLEMENTS

University of Warwick

ANA BEATRIZ C. GALVÃO

European University Institute

We test the expectations theory of the term structure of U.S. interest rates in nonlinear systems. These models allow the response of the change in short rates to past values of the spread to depend upon the level of the spread. The nonlinear system is tested against a linear system, and the results of testing the expectations theory in both models are contrasted. We find that the results of tests of the implications of the expectations theory depend on the size and sign of the spread. The long maturity spread predicts future changes of the short rate only when it is high.

Keywords: Term Structure, Expectations Theory, Threshold Models

1. INTRODUCTION

A number of recent studies of the term structure of U.S. interest rates look for possible asymmetries in the response of the short-term interest rate to spreads between long and short rates; earlier work has tested the expectations theory of the term structure by imposing restrictions in vector autoregressive (VAR) models, following Campbell and Shiller (1987), or by regressions of the change in short rates on the lagged spread [see, e.g., Mankiw and Miron (1986)]. The work on asymmetries in the response of the short-term interest rate has generally been based on single-equation threshold models for the short rate. We show that single-equation nonlinear threshold model testing and specification techniques can be adapted to permit joint modeling of the short-term interest rate and the spread, and that some of the tests proposed by Campbell and Shiller (1987) can be applied in such a framework. The systems approach yields interesting insights into the

We acknowledge the helpful comments of two anonymous referees. Financial support from the UK Economic and Social Research Council under grant L138251009 is gratefully acknowledged by the first author, and from Capes-Brazil by the second author. The computations reported in this paper were performed using code written in the Gauss Programming Language. Address correspondence to: Michael P. Clements, Department of Economics, University of Warwick, Coventry CV4 7AL UK; e-mail: M.P.Clements@warwick.ac.uk.

dynamic behavior of the variables, and we are able to interpret the estimated models in terms of recent work on the term structure of interest rates.

Sola and Driffill (1994) are perhaps closest to our approach, in that expectations theory restrictions are applied in a bivariate VAR of 3- and 6-month interest rates in which the means and equation error variances are permitted to follow an unobserved Markov process. Matching Hamilton (1988), they find that the 1979–1982 period, when the Federal Reserve changed its operating procedures, is associated with a different regime than the periods 1962–1978 and 1983–1990. Furthermore, the expectations theory restrictions are not rejected in the Markov-switching model, although they are in a constant-parameter, linear VAR. In addition to 3- and 6-month rates, we consider 3-month and 10-year interest rates, extend the sample to include the 1990's, and allow up to three regimes, with the regime in force at any period being determined by the (lagged) value of the spread, rather than being exogenously given, as in the Markov-switching approach.

The dependence of the regime on the spread is a potentially attractive feature because it explicitly models the idea that adjustment to equilibrium in financial markets may depend on the sign of the disequilibria and may not vary proportionately with its size. In that case, linear equilibrium-correction models will be inappropriate. The frequently cited example is of transactions costs, whereby arbitrage opportunities between two markets only arise when the price differential is large enough to imply net gains to traders. As a consequence, nonlinear equilibrium correction models have been used to model the relationship between spot and future prices by, *inter alia*, Dwyer et al. (1996), Martens et al. (1998), and Tsay (1998), and between interest rates of different maturities by, for example, Anderson (1997), Enders and Granger (1998), van Dijk and Franses (2000), and Hansen and Seo (2001).

The plan of the remainder of this paper is as follows: Section 2 reviews vector equilibrium-correction models (VECM's) of the term structure, and VAR-based tests of the expectations theory. The specification and testing procedures for threshold VECM models are recorded in Section 3, and the empirical results are presented in Section 4. The relative forecast performance of a number of linear and nonlinear models of interest rates are considered elsewhere [Clements and Galvão (2001)]. Section 5 draws out the economic implications of our results and relates these to the recent literature. Section 6 concludes.

2. TERM STRUCTURE OF INTEREST RATES AND EQUILIBRIUM CORRECTION MODELS

The simple expectations theory implies that the k -period interest rate is the weighted average of the expected future one-period interest rates plus a risk premium:

$$r_t(k) = \frac{1}{k} \left[\sum_{j=1}^k E_t r_{t+j-1}(1) \right] + L_t(k), \quad (1)$$

where $r_t(s)$ is the s -period interest rate at t , E_t is the expectations operator conditional on time t , and $L_t(k)$ is the term premium, which may reflect risk and liquidity premia. Arbitrage between bond markets with different maturities will ensure that this condition holds, while the presence of the term premium will generally result in the yield curve [$r_t(k)$ plotted against k] being upward sloping. If $r_t(1)$ is integrated of order one [$I(1)$], then so is $r_t(k)$ from (1), and interest rate spreads are $I(0)$, because (1) can be written as

$$S_t(k) \equiv r_t(k) - r_t(1) = \frac{1}{k} \left[\sum_{i=1}^{k-1} \sum_{j=1}^i E_t \Delta r_{t+j}(1) \right] + L_t(k) \quad (2)$$

$$= \frac{1}{k} \sum_{j=1}^{k-1} (k-j) E_t \Delta r_{t+j}(1) + L_t(k) \quad (3)$$

[see, e.g., Hall et al. (1992)]. The RHS is the sum of a finite number of $I(0)$ terms (ignoring the risk premium) and is therefore $I(0)$. Thus, $r_t(k)$ and $r_t(1)$ are cointegrated with weights $[1 - 1]$, as are any two yields of different maturities. By the Granger representation theorem, cointegration implies the existence of a VECM, viz.

$$\Delta \mathbf{r}_t = \Phi(L) \Delta \mathbf{r}_{t-1} + \alpha (S_{t-1} - \mu) + \varepsilon_t, \quad (4)$$

where $\mathbf{r}_t = [r_t(l), r_t(s)]'$ denotes the vector of the long ($k = l$) and short-term rates ($s = 1$); $\Phi(L)$ is a matrix of coefficients in the lag operator, L , with $L^n \mathbf{x}_t = \mathbf{x}_{t-n}$, and $\Delta = 1 - L$ the difference operator. S_t is the spread and μ is the equilibrium spread; they may differ from zero because of the term premium. The adjustment vector to the long-run attractor is α .

2.1. Expectations Theory

To make matters concrete, we suppose that $\Phi(L)$ is first order, so that we have lag 1 and lag 2 terms in the changes in $r_t(l)$ and $r_t(s)$:

$$\Delta \mathbf{r}_t = \mathbf{c} + \sum_{j=1}^2 \Phi_j \Delta \mathbf{r}_{t-j} + \alpha \beta' r_{t-1} + \varepsilon_t, \quad (5)$$

where $\beta' = [1 : -1]$, so that $\beta' \mathbf{r}_{t-1} = S_{t-1}$ defines the spread. As Campbell and Shiller (1987) note, because by equation (2) the spread is a linear combination of future changes in the short rate [plus $L_t(k)$], it should help predict the values of future changes in those rates. This can be formulated as the hypothesis that S Granger causes $\Delta r(s)$, which can be tested within a single-equation autoregressive distributed-lag model of $\Delta r(s)$ on S . However, for subsequent tests, it will be convenient to specify a VAR in S and $\Delta r(s)$. This can be obtained directly by

premultiplying (5) by B , where

$$B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

to give

$$B\Delta r_t = Bc + \sum_{j=1}^2 B\Phi_j B^{-1}(B\Delta r_{t-j}) + B\alpha S_{t-1} + B\varepsilon_t, \quad (6)$$

where $B\mathbf{r}_t = [S_t \ r_t(s)]'$. When α is not the zero vector (so that the two $I(1)$ interest rates are cointegrated with vector $[1: -1]$), equation (6) can be rearranged to give a VAR in $\mathbf{r}_t^* = [S_t \ \Delta r_t(s)]'$, with $\mathbf{v}_t = B\varepsilon_t$, $\mathbf{c}^* = Bc$, and suitably defined coefficient matrices¹:

$$\mathbf{r}_t^* = \mathbf{c}^* + \Gamma_1 \mathbf{r}_{t-1}^* + \Gamma_2 \mathbf{r}_{t-2}^* + \Gamma_3 \mathbf{r}_{t-3}^* + \mathbf{v}_t. \quad (7)$$

If, however, $\alpha = \mathbf{0}$, the levels term in the interest-rate spread is absent from (6), so that S is integrated of order 1, and the VAR corresponding to (7) is in $[\Delta S_t \ \Delta r_t(s)]$, so that there are restrictions on the Γ_i in (7) such that S has a unit root [and, in addition, $\Gamma_3 = \mathbf{0}$]. Therefore, the VAR is constructed on the assumption that the theoretical implication of the expectations theory—that $r(l)$ and $r(s)$ are cointegrated—holds.

Because the expectations theory posits a relationship between the spread and predictions of future changes in $r(s)$, it will prove useful to write (7) in companion form, whence the predictions have a relatively simple analytical form. The companion form is

$$\mathbf{R}_t = \mathbf{C} + \Gamma \mathbf{R}_{t-1} + \mathbf{V}_t, \quad (8)$$

where

$$\mathbf{R}_t = \begin{bmatrix} \mathbf{r}_t^* \\ \mathbf{r}_{t-1}^* \end{bmatrix}, \mathbf{C} = \begin{bmatrix} \mathbf{c}^* \\ \mathbf{0} \end{bmatrix}, \Gamma = \begin{bmatrix} \Gamma_1 & \Gamma_2 \\ I_2 & \mathbf{0} \end{bmatrix}, \mathbf{V}_t = \begin{bmatrix} \mathbf{v}_t \\ \mathbf{0} \end{bmatrix}.$$

We define $\mathbf{i}_S = [1 \ 0 \ 0 \ 0]$ and $\mathbf{i}_r = [0 \ 1 \ 0 \ 0]$, so that $S_t = \mathbf{i}_S \mathbf{R}_t$ and $\Delta r_t(s) = \mathbf{i}_r \mathbf{R}_t$.

At this point, we note two versions of the expectations theory. Underlying the stronger version of the theory is that $L_t(k) = \theta$ in (2), so that the term premium is constant. Then, taking expectations dated period t of both sides of (2) gives

$$E_t S_t \equiv S_t = \frac{1}{k} \left[\sum_{j=1}^{k-1} (k-j) E_t \Delta r_{t+j}(1) \right] + \theta. \quad (9)$$

As Sola and Driffill (1994, p. 604) note, it is common to posit a slightly weaker version that allows for a random error term in the relationship between long- and short-term rates, so that $L_t(k) = \theta + u_t$, where θ is again the constant term premium and u_t represents measurement error or random variation in the premium. The random element has the key property that it is an innovation with respect to

expectations taken at period $t - 1$, so that, from (2),

$$E_{t-1}S_t = \frac{1}{k} \left[\sum_{j=1}^{k-1} (k-j) E_{t-1} \Delta r_{t+j}(1) \right] + \theta, \quad (10)$$

where we use $E_{t-1}E_t(\cdot) = E_{t-1}(\cdot)$. We entertain both possibilities, noting that a key attraction of (9) is that it can be applied directly in the context of threshold models, at least for $k = 2$, as explained in Section 2.2.

2.2. Cross-Equation Restrictions in the VAR

Beginning with (10), the LHS of this expression from (8) is $i_s C + i_s \Gamma R_{t-1}$. To evaluate the RHS of (10), note that

$$E_{t-1} \Delta r_{t+j}(1) = i_r E_{t-1} R_{t+j} = i_r \left[C \sum_{s=0}^j \Gamma^s + \Gamma^{j+1} R_{t-1} \right].$$

In Appendix A, we derive the following approximate expression for the term in R_{t-1} for the RHS of (10) (omitting θ), for large k :

$$i_r \Gamma^2 (I - \Gamma)^{-1} R_{t-1}.$$

By equating coefficients on S_{t-1} on both sides of (10) and rearranging, we obtain

$$i_s \Gamma (I - \Gamma) = i_r \Gamma^2. \quad (11)$$

When $k = 2$, equation (10) simplifies to $E_{t-1}S_t = \frac{1}{2}[E_{t-1}\Delta r_{t+1}(1)] + \theta$, so that substituting from the VAR gives

$$i_s C + i_s \Gamma R_{t-1} = \frac{1}{2} i_r \left[C \sum_{s=0}^1 \Gamma^s + \Gamma^2 R_{t-1} \right] + \theta. \quad (12)$$

Equating coefficients on R_{t-1} again gives four cross-equation restrictions on the coefficients in the VAR:

$$i_s \Gamma = \frac{1}{2} i_r \Gamma^2. \quad (13)$$

The restrictions for $k = 2$ and the large k case in terms of the elements of the matrices of the VAR are detailed in Appendix B.

For $k = 2$, the stronger version of the theory implies $i_s R_t \equiv S_t = \frac{1}{2} i_r E_t[R_{t+1}] + \theta = \frac{1}{2} i_r [C + \Gamma R_t] + \theta$, so that the restrictions are (equating coefficients on R_t)

$$i_s = \frac{1}{2} i_r \Gamma$$

and, for large k ,

$$i_s (I - \Gamma) = i_r \Gamma. \quad (14)$$

These restrictions are also detailed in Appendix B.

For $k = 2$ and the stronger version of the theory, we can readily calculate the “theoretical” spread from the VAR because the one-step-ahead predictions of $\Delta r_{t+1}(s)$ are simply the fitted values in the VAR equation for $\Delta r_{t+1}(s)$ (we replace the unknown parameters in the VAR with their full-sample estimates of their values). Therefore, the theoretical spread (S') is simply one-half times the fitted value. More generally, as in Campbell and Shiller (1987, p. 1080, Fig. 1), plots of S_t and S'_t under the expectations theory can be obtained using the VAR-based predictions of the short rate to evaluate the RHS of (9) to give S'_t .

In the case of large k , the long-term bond carries coupons, which have a higher present value in the near future than in the distant one. Therefore, S'_t can be modified accordingly to take into account weights declining monotonically with the time horizon j :

$$\frac{1}{1 - g^k} \sum_{j=1}^{k-1} (g^j - g^k) i_r E_t[\mathbf{R}_{t+j}], \quad (15)$$

where g is an approximation for returns on bonds selling close to par, that is, $g = 1/[1 + \mu_{r(l)}]$, given that $\mu_{r(l)}$ is the average of the long-term rate [see Hardouvelis (1994)]. These weights are based on a linearized approximation to the yields on coupon bonds and $(1 - g^k)/(1 - g)$ is the “duration” of a k -period bond.

2.3. Granger Causality of the Short Rate by the Spread

The weaker restriction implied by the expectations theory, that S does not Granger-cause $\Delta r(s)$, can be tested by $\Gamma_{i,21} = 0$, $i = 1, 2, 3$, where, for example, $\Gamma_{1,21}$ is the coefficient on S_{t-1} in the equation for $\Delta r_t(s)$.

3. THRESHOLD MODELS

The empirical evidence of cointegration between yields of different maturities, and the ability of the ECM to improve forecasts of interest rates, is contested in the literature [see, e.g., Pagan et al. (1996)]. The ability of S to forecast $r(s)$ is found to depend on the maturities [Rudebusch (1995)] and on the monetary policy in operation [Mankiw and Miron (1986), Rudebusch (1995), Roberds et al. (1996), and Gray (1996)]. Tzavalis and Wickens (1998) note that the presence of the time-varying risk premium in (4) may mask the predictive power of the spread, and the dynamics of $\Delta r(s)$ may depend on its level, as in the regime-switching models considered by Pfann et al. (1996). Moreover, the expectations theory ignores transaction cost effects. These considerations—the diversity of monetary policies, the term premium, and the presence of transaction costs—suggest nonlinear equilibrium correction models, which allow the speed of adjustment to equilibrium to depend on the regime, may be warranted.

Allowing for asymmetric adjustment to positive and negative values of past realizations of S_t , or to changes in S_t , motivates the threshold autoregressive (TAR) and momentum TAR (MTAR) models of Enders and Granger (1998). Anderson (1997) allows for no adjustment to equilibrium when S_t lies within a band around μ (where the band will only be centered on μ when transaction costs are symmetric), and for adjustment at different speeds outside the band. Heterogeneity in the transaction costs faced by individuals implies that in the aggregate the effect of transaction costs might be better modeled by a “smooth” transition equilibrium correction model (STECM): see Anderson (1997) and van Dijk and Franses (2000). These are single-equation analyses, although van Dijk (1999, Ch. 5, p. 128) discusses bivariate STVEC (smooth transition vector equilibrium correction) models.

Nonlinearity is modeled with threshold vector equilibrium correction models (TVECM), which allow the coefficients on the dynamics and the equilibrium correction term to alter from one regime to another:

$$\Delta \mathbf{r}_t = \mathbf{c}_i + \sum_{j=1}^2 \Phi_{i,j} \Delta \mathbf{r}_{t-j} + \alpha_i S_{t-1} + \varepsilon_t, \quad (16)$$

where $\Phi_{i,j} = \Phi_{s,j}$ and $\alpha_i = \alpha_s$ if $\gamma_{s-1} < z_{t-d} \leq \gamma_s$, for $s = 1, 2$ in the case of models with two regimes and $s = 1, 2, 3$ for a three-regime model. The transition variable is taken to be the spread, $z_{t-d} = S_{t-d}$, with delay $d = 1$ [following, e.g., Anderson (1997) and Hansen and Seo (2001)]; the thresholds are defined by γ_i (with $\gamma_0 = -\infty$, $\gamma_3 = +\infty$). This model differs from the switching-regime model of Sola and Driffill (1994) because the regimes are given by the equilibrium correction term—the spread—and not by an exogenous unobserved variable.

Equation (16) implies a threshold VAR (TVAR) for \mathbf{r}_t^* ($= [S_t \ \Delta \mathbf{r}_t(s)]'$) with identical values of the thresholds $[\gamma_1 \ \gamma_2]$. This is because \mathbf{r}_t^* is a nonsingular linear transformation of $\Delta \mathbf{r}_t$, so that the values of the thresholds that minimize the determinant of the system error covariance matrix for \mathbf{r}_t^* and $\Delta \mathbf{r}_t$ will be the same.

3.1. Testing Whether the Spread Granger-Causes the Short Rate

Writing the TVAR in \mathbf{r}_t^* as

$$\mathbf{r}_t^* = \mathbf{c}_i^* + \sum_{j=1}^2 \Gamma_{i,j} \mathbf{r}_{t-j}^* + \mathbf{v}_t, \quad (17)$$

the test of Granger noncausality (GNC) of S for $\Delta \mathbf{r}(s)$ is that $[\Gamma_{i,j}]_{21} = 0$ for all i and $j = 1, 2$, where $[\Gamma_{i,j}]_{21}$ denotes the $(2, 1)$ element of the lag j matrix of coefficients in regime i . To investigate the impact on the testing procedure of conditioning on the estimated regimes, when the threshold effects may not be especially strong, we simulate the empirical distribution of the test, allowing for the

uncertainty inherent in the estimation of the thresholds. The estimated TVAR—with the null of GNC imposed—is taken as the data-generating process (DGP) and used to simulate data by resampling with replacement (i.e., bootstrapping) the residuals. We then calculate the GC tests, conditioning on the true (i.e., DGP) values of the thresholds and after selecting the thresholds on the simulated data. To calculate the GC tests, we estimate restricted (GNC imposed) and unrestricted models. In the case in which the thresholds are estimated, the values in the restricted models are set equal to those estimated in the unrestricted models (mirroring the construction of the empirical value of the GC test). By repeating this exercise a number of times, we estimate the empirical null distributions of the GC test both when we condition on the thresholds and when their sampling variability is taken into account, and so obtain an assessment of the impact on the GC-test null distribution of the conditioning.

3.2. Tests of Cross-Equation Restrictions

The problem with TVAR models is that closed-form expressions for multistep expectations are not available.² However, for testing the strong form of the expectations theory, when $k=2$, only one-step predictions are required. We can write (17) for the case of two regimes (for notational simplicity only) as

$$r_t^* = [c_1^* + d^*] \begin{bmatrix} 1 \\ \mathbf{1}(s_t) \end{bmatrix} + \sum_{j=1}^2 [\Phi_{1,j}^* : \Theta_j] \begin{bmatrix} r_{t-j}^* \\ r_{t-j}^* \times \mathbf{1}(s_t) \end{bmatrix} + v_t, \quad (18)$$

where $\mathbf{1}(s_t) = 1$ when s_t is “true,” and s_t is defined as the event that $S_{t-1} > r$. The equivalence of (17) and (18) is apparent from $c_1^* + d^* = c_2^*$, and $\Phi_{1,j}^* + \Theta_j = \Phi_{2,j}^*$, $j=1, 2$. Then, the strong condition implies that $S_t = \frac{1}{2} i_2 E_t[r_{t+1}^*] + \theta$, where $i_2 = [0 \ 1]$, so that $i_2 E_t[r_{t+1}^*]$ is just the one-step predicted value of $\Delta r_{t+1}(s)$ made at time t . This depends on the value of s_{t+1} , and thus S_t , which is known at t . So, when the unknown parameters are replaced with full-sample estimates, the one-step predicted values of $\Delta r_{t+1}(s)$ at time t are just the fitted values for the $t+1$ observations.

For k in excess of two, we require terms such as $E_t[\Delta r_{t+j}(s)]$ for $j > 1$, for which s_{t+j} is not known. As shown by, for example, Granger and Teräsvirta (1993) and Clements and Smith (1997), these expectations can be evaluated in a number of ways, including Monte Carlo or bootstrapping. Thus, conditional on data up to period t (but on full-sample parameter estimates), the j -step-ahead conditional prediction density can be constructed by iterating the model forward with future disturbances replaced with either pseudorandom numbers or bootstrapped residuals. The distribution of the residuals is taken conditionally on the regime. The mean of this density is an estimator of the conditional expectation. In this way, we can evaluate the sum of expected future changes in $r(s)$ over the relevant horizons, given k and conditional upon the TVAR, and compare this with the actual spread, S_t . For large k , the theoretical spread takes

into account coupons on the long bond and uses the decreasing weights given in equation (9).

3.3. Estimation and Testing of Threshold VAR Models

Conditional on the thresholds and the transition variable, a TVAR model can be estimated by multivariate least squares. The value of the thresholds (γ) is estimated by a grid search. For each possible value of γ (a scalar for the two-regime model, a two-element vector for the three-regime model, and restricted such that at least 10% of the observations fall in each regime), we calculate $\ln |\Omega(\gamma)|$, where $\Omega(\gamma) = [\epsilon(\gamma)' \epsilon(\gamma)]/T$ is the estimated covariance matrix of the residuals for a particular value of γ . The estimator of γ is the value that minimizes the log determinant of Ω .³ An alternative approach due to Balke and Fomby (1997) is to estimate γ from a univariate threshold autoregression for the spread, with the spread lagged one period as the transition variable.

Nonlinearity can be tested by comparing the value of the maximized loglikelihood of the VAR with the TVAR. If we denote the estimated covariance matrix of the linear system by $\hat{\Omega}_1$, that of two- and three-regime homoskedastic models by $\hat{\Omega}_2$ and $\hat{\Omega}_3$, and let T be the effective number of observations, then the LR test,

$$LR_{i,i+1} = T \{ \ln[\det(\hat{\Omega}_i)] - \ln[\det(\hat{\Omega}_{i+1})] \}, \quad (19)$$

compares the linear model against a two-regime model when $i = 1$, and the two-regime model against the three-regime model when $i = 2$ (a direct comparison of the three-regime model against linearity is also possible). The asymptotic distribution is an extension of Hansen (1996), as argued by Hansen and Seo (2001), and the bootstrap can be used to obtain a finite sample approximation. For $i = 1$, the bootstrap distribution is calculated from data generated by the linear model by resampling its residuals, where the residuals are corrected for heteroskedasticity using a regression of the squared residuals on the squared regressors, as described in Hansen (2000).

4. EMPIRICAL RESULTS

Two samples with different long-rate maturities of quarterly frequency are employed in the evaluation of the implications of the expectations theory: 3- and 6-month Treasury Bill interest rates ($k = 2$) and 3-month bill and 10-year Treasury bond interest rates ($k = 40$). Monthly observations on these series are taken from the Federal Reserve Bank of St. Louis database (Fred: www.stls.frb.org/fred), and are plotted in Figure 1. For the 3-month bill and the 10-year Treasury bond interest rates, there are observations from 1953:04 to 2001:10; for the 6-month rate the observations begin in 1958:12. The first panel shows the time series of the 3-month and 10-year rates, the middle panel the spread between these rates, and the bottom panel the spread between the 6- and 3-month rates. The quarterly observations used in the subsequent analysis are the rates for the last month in the quarter.

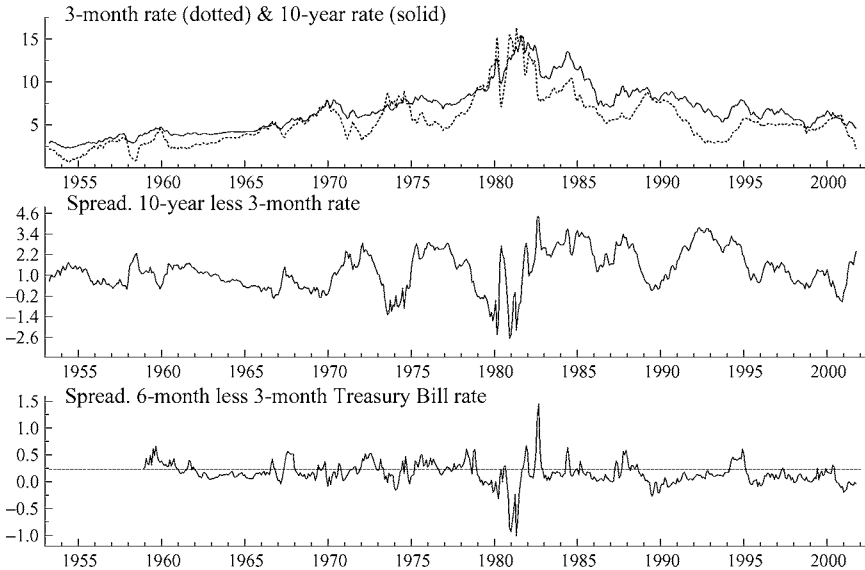


FIGURE 1. Time series of monthly interest rates and spreads.

4.1. Results Based on Linear VAR

Table 1 records the results of testing the expectations hypothesis in a linear framework for both data sets. Neither the Granger causality tests nor the cross-equation restrictions provide any support for the expectations hypothesis on the 3- and 6-month data set—GNC of the spread for the short rate is not rejected, and the cross-equation restrictions are rejected at the 1% level. For 3-month and 10-year interest rates, the rejection of the cross-equation restrictions is a good deal less pronounced, and occurs at approximately the 10% level for both the strong and the weak versions of the theory. Figure 2 shows the cross plots between the theoretical spread and the actual spread for both data sets (top left panel for the 3- and 6-month rates, middle left for the 3-month and 10-year rates). Both curves are flatter than the 45-degree line, with the interpretation that the theoretical spread (S') underpredicts the actual (S). The solid lines in the cross plots in Figure 2 are estimated using local linear regression in the SPLUS program. Specifically, we employ *loess* (which is robust to boundary effects and outliers) with a span of 0.3. S' is the weighted cumulative sum of the forecasts for the short-term rate using the VAR coefficients for the full sample and data information until t for $(k - 1)$ -steps ahead [see equation (15)]. In all of the calculations with $k = 40$, we set $g = 0.937$, given that the mean of the 10-year rate is 6.71. Thus the curves are calculated under the stronger version of the hypothesis. In the case of $k = 2$, the theoretical spread is the fitted value of the short-term rate multiplied by $1/2$ at each t . Table 1 also records the correlation between S and S' , the slope (ρ) as another measure of linear dependence, and their standard deviations, for comparison with the same statistics for the nonlinear models.

TABLE 1. Testing the expectations hypothesis using VAR's^a

Model/test	Results
A. 3- and 6-month rates ($s = 3$ and $l = 6$)	
Linear model	
Noncausality tests of S for $\Delta r(s)$	3.53[0.060]
Test of term structure restrictions	
weak	$\chi^2(2) = 11.16[0.004]$
strong	$\chi^2(2) = 11.38[0.003]$
$\rho_{S',S}$	0.357
$\sigma_{S'}/\sigma_S$	0.552
β	0.197
Linearity tests	
VAR vs. 2R-TVAR	68.35[0.007]
VAR vs. 3R-TVAR	92.07[0.036]
2R-TVAR vs. 3R-TVAR	23.73[0.595]
Nonlinear model with two regimes ($\gamma = 0.23$)	
Noncausality tests S for $\Delta r(s)$	
1st regime	0.439(0.520)[0.570]
2nd regime	0.832(0.356)[0.402]
All	1.269(0.128)[0.128]
$\rho_{S',S}$	0.189
$\sigma_{S'}/\sigma_S$	0.884
β	0.315
B. 3-month and 10-year rates ($s = 3$ and $l = 120$)	
Linear model	
Noncausality tests of S for $\Delta r(s)$	3.01[0.083]
Test of term structure restrictions	
weak	$\chi^2(2) = 5.27[0.072]$
strong	$\chi^2(2) = 4.67[0.097]$
$\rho_{S',S}$	0.760
$\sigma_{S'}/\sigma_S$	0.437
β	0.332
Linearity tests	
VAR vs. 2R-TVAR	56.27[0.018]
VAR vs. 3R-TVAR	77.52[<0.001]
2R-TVAR vs. 3R-TVAR	133.79[<0.001]
Nonlinear model with three regimes ($\gamma_1 = -0.07$ and $\gamma_2 = 2.6$)	
Noncausality tests of S for $\Delta r(s)$	
1st regime	13.53(0.401)[0.359]
2nd regime	2.85(0.266)[0.308]
3rd regime	14.23(<0.000)[0.003]
All	29.31(0.119)[0.164]

(continued)

TABLE 1. (Continued.)

Model/test	Results
$\rho_{S',S}$	0.639
$\sigma_{S'}/\sigma_S$	0.766
β	0.489

^a S' is the theoretical spread and S is the actual; ρ is the correlation between S and S' , and σ_S and $\sigma_{S'}$ are their standard deviations; $\beta = \rho\sigma_{S'}/\sigma_S$ is the slope. The lag orders of the VAR and TVAR models (selected by SIC) are 1 for both data sets. Some experimentation for the VAR's indicated that the GNC and cross-equation restriction tests were qualitatively unaffected by setting $p = 5$. The p -values of the GNC tests for the TVAR's are calculated using a bootstrap as described in the text: The p -values given in parenthesis condition on the empirical estimates of the thresholds, whereas those in square brackets permit regime uncertainty.

In summary, the results based on the VAR's do not support the expectations theory. They match the findings of Hardouvelis (1994) on a similar data set but differ from the results of Campbell and Shiller (1987) for different frequencies and maturities.

4.2. Results Based on TVAR

Table 1 also records the results based on the two- and three-regime TVAR models, as well as tests of the TVARs against the VAR. For both data sets the linear VAR is rejected in favour of the non-linear TVAR.

4.2.1. *Three- and six-month interest rates.* The tests of the VAR against the 2R-TVAR and the 3R-TVAR both reject at the 5% level, and the specification test of two versus three regimes using heteroskedasticity-corrected p -values finds in favor of the 2R-TVAR. We find $\hat{\gamma} = 0.23$, placing 29% of the 169 observations in the upper regime. The estimates of the probabilities of remaining in the same regime, and of switching regimes, are

$$\begin{bmatrix} 0.800 & 0.200 \\ 0.510 & 0.490 \end{bmatrix},$$

where the (i, j) element is the probability of being in regime j at time t given that regime i was operative at $t - 1$. Thus, there is a more or less equal chance of staying in the upper regime, and switching from the upper to the lower. The tests for GNC fail to reject the null hypothesis in either regime. The correlation (ρ) between S and S' decreases relative to the linear model, but the slope coefficient (β) increases because $\sigma_{S'}$ better matches σ_S . This is readily apparent in Figure 2, where comparing the top right panel to the top left one shows the increased dispersion in S' from allowing nonlinearity. Otherwise, allowing for nonlinearity has little effect, and the cross plot is no closer to the 45-degree line.

Allowing for nonlinearity does little to rescue the expectations theory, in contrast to findings by Sola and Driffill (1994). They found that the 1979–1982 episode of high and variable interest rates is picked out as belonging to a separate regime.

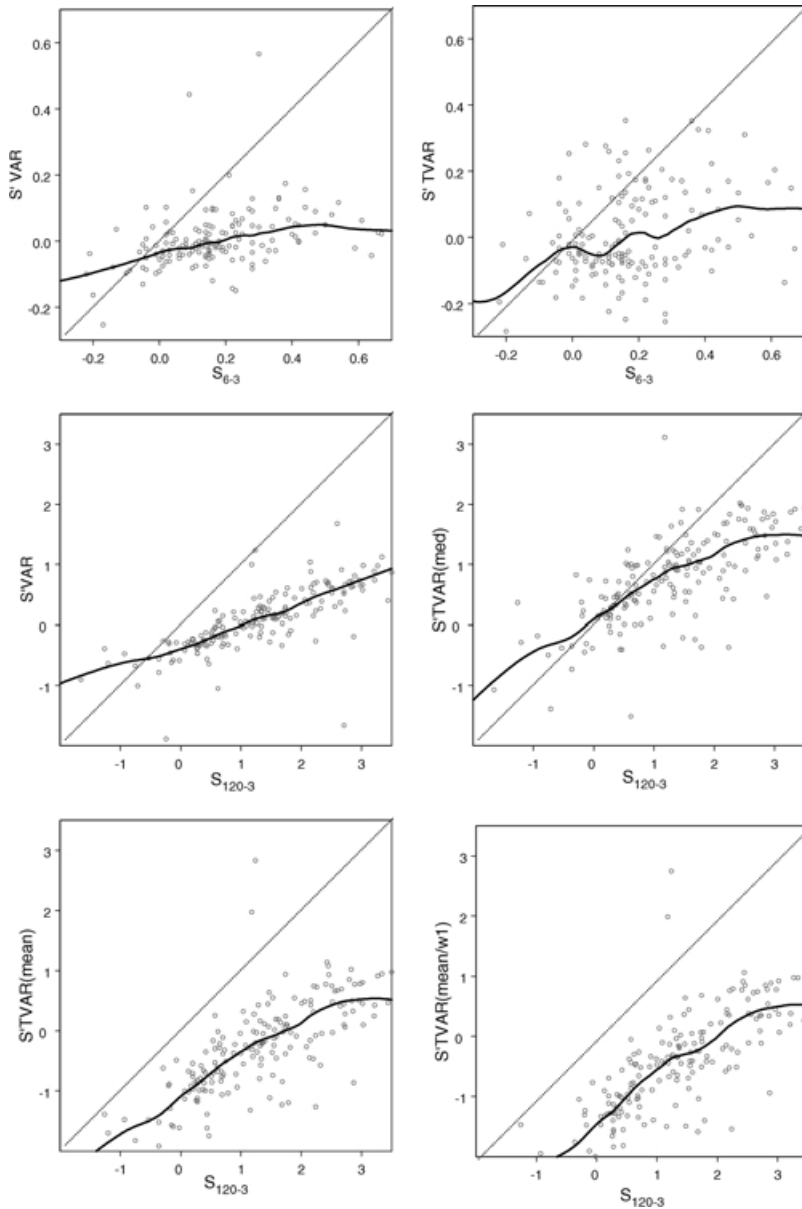


FIGURE 2. Predicted S' and actual values S of the spread: (upper panel) 6- and 3-month rates using VAR and 2R-TVAR models; (middle panel) 120- and 3-month rates using VAR and 3R-TVAR models, with the point forecasts taken to be the medians of the forecast densities; (lower panel) same as middle panel except point forecasts given by the means of the forecast densities, and with the weights used in the construction of S' as in equation (15) (left) or equation (3) (right).

Our 2R-TVAR does not entail a similar partitioning of the data, and it seems that there is no short-rate dependence on the size of the spread for this short maturity. The lower panel of Figure 1 plots S_t with a horizontal line drawn at the estimated threshold value. It is apparent that observations in the 1979–1982 period fall in both regimes.

4.2.2. Three-month and ten-year interest rates. The model specification tests suggest a 3R-TVAR, and that the dynamics of the short rate depend on the sign and size of the spread. The estimated thresholds are -0.07 and 2.6 , which places 10% of the observations (19) in the lower regime and 15% in the upper. The empirical transition probabilities are

$$\begin{bmatrix} 0.684 & 0.263 & 0.053 \\ 0.042 & 0.889 & 0.069 \\ 0 & 0.379 & 0.621 \end{bmatrix},$$

implying a longer duration of the middle regime, and an absence of shifts from large spreads (>2.60) to negative ones (<-0.07). The standard deviation of the residuals of the short-rate equation is four times larger in the lower spread regime, at 2.08, than in the upper regime (0.51), while in the middle regime the value is around 1. Although the coefficient of the lagged spread in the short-rate-changes equation is largest (estimated at around 1) in the lower regime, it is not statistically significant, given the large degree of uncertainty that characterizes this regime, as confirmed by the GNC test results presented in Table 1. GNC is only rejected in the upper regime, which is the regime with the smallest residual variance.

The predictions of the changes in the short rate are generated from the 3R-TVAR [to calculate S' as in equation (9)] using 1,000 sequences of $k - 1$ (i.e., 39) forecasts based on bootstrap samples of the residuals. We condition on the full-sample estimates of the coefficients, and use data information up to $t - 1$. We draw from the sub-sample of the full-sample estimated residuals to form the bootstrap samples, where the sub-sample consists of the set of regime-specific residuals warranted by the regime the model is in at that point. We calculate the point forecasts of the changes in the short rate for each one of the $k - 1$ horizons as both the mean and median of the simulated forecast densities. Strictly, the mean is to be preferred because it is the conditional expectation, but the median is more robust to the possibility that a small number of the replications might produce outliers.

The values of ρ , β , and the standard deviations σ_S and $\sigma_{S'}$ recorded in Table 1 are calculated using point forecasts estimated as the means of the forecast distributions. The middle right panel of Figure 2 shows the cross plot of S and S' using the median, and the bottom left panel shows the same plot using the mean. We find that S' is similar to the VAR calculations within the middle-regime boundaries when the means are used, but that the outer regime behavior is different. For negative values of S , there appears to be no relation between S' and S in the linear model, but there is a positive relationship in the TVAR. For large values of the spread, the

positive relationship is flatter than for the VAR. Comparing results of calculating S' using the means and medians, we find that, for the mean, negative future changes are predicted for intermediate values of the spread, whereas for the median, the predicted changes are positive. The shift upward of the cross plot using the median constitutes stronger evidence in favor of the expectations hypothesis.

We also check the sensitivity of these results to the assumption of a coupon effect. Using the weights without the coupons effect [equation (3)], and the mean of the forecast distribution to obtain point forecasts for each horizon, we calculated S' : see Figure 2, bottom right panel. There is a slight shift downward of the cross plot compared to the bottom left panel (with coupon weights), but little substantive difference is noticeable. We also checked the sensitivity of the results to different choices of p (we have hitherto reported results for the SIC-selected $p = 1$ for the VAR's and TVAR's), but little difference was observed in the cross plot.

In summary, stronger evidence for the expectations theory results from allowing for non-linearities, especially when S is relatively large.

5. ECONOMIC ANALYSIS OF THE MODELS

There is a large academic literature on the term structure of interest rates. In this section, we discuss our findings in the light of some of the recent contributions, focusing on the long-maturity-spread case. The implication of the 3R-TVAR is that the spread between the 10-year and the 3-month rates only predicts future changes in the short rate when it is large and positive. We estimated a 3R-TVECM [equation (16)] for this data set (not reported), which indicated that the spread also predicts *decreasing* long rates in the high, positive spread regime, which is at odds with the expectations theory. For example, Campbell (1995, pp. 137–138) found that “[w]hen long rates are unusually high relative to short rates, long rates do not decline to restore the usual yield curve, as one might suppose. Instead long rates tend to rise; the yield spread falls only because short rates rise even faster.”

Thus for high positive values of the spread, our findings are consistent with the puzzle presented in the literature that “the movement of future cumulative short rates obey[s] the overall direction predicted by the expectations hypothesis but at the same time the short-run movement of the long rates does not” [Hardouvelis (1994, p. 256)]. However, for lower values of the spread, future cumulative short rates do not appear to behave as expected either.

Relative to the VAR, however, the 3R-TVAR is able to characterize the responses that occur after possible overreactions (underreactions) of the long-term interest rate to changes in monetary policy. From Figure 1, it is apparent that the upper regime is more prevalent after 1984. This may indicate a change in the stance of monetary policy to a more aggressive smoothing of the short rate by the FED. The credibility of the monetary authority appears to have been enhanced by the pursuance of persistent and well-defined policies and, as Watson (1999), argues, even small increases in the persistence of the short-term interest rate, especially at

already high levels of persistence, would, according to standard models, translate into substantially higher volatility in the long-term rate, a feature of the second half of the 1980's and 1990's. Then, small changes in short rates can engineer large changes in long rates, as noted by Campbell (1995, pp. 142–145) in his discussion of the impact of monetary policy on the bond market in the spring of 1994. One strand of argument is that the increased variability of long rates leads to their overreaction to monetary authority policies aimed at preventing the economy from overheating, so that in subsequent periods the disequilibrium between the expected values of the short-term rate and the actual long-term rate leads to decreasing long rates, establishing a negative correlation between the spread and the long rates.⁴

6. CONCLUSIONS

We show how a threshold VAR of the spread between long- and short-term interest rates and the changes in the short rate arise from a threshold VECM for long and short rates, where the spread is the equilibrium correction term. The threshold nonlinearities can be viewed as resulting from transaction costs, changes in policy regime, time-varying risk premia, etc., and suggest that whether or not the restrictions imposed by the expectations theory may depend on the value of the spread. We consider Granger causality tests of the spread for future changes in the short-rate in the TVAR, and whether allowing for nonlinearity gives a closer match between the actual and theoretical spreads, where the latter is the cumulation of model-based future predicted changes in the short rate.

We show that the dynamics of the short rate and the spread depend on the value of the spread, and that this nonlinear dependence is strong enough to change the results of tests of the expectations theory when the long rate is taken as the 10-year bond interest rate. Then, we find that short rates behave in accordance with the theory when the spread is large and positive. Even then, however, the spread also predicts decreasing long rates. This puzzle has been reported in the literature, but the TVAR model shows that whether the responses to the spread are significant depends on the size of the spread.

NOTES

1. In fact, Γ_3 has zeros in the last column, so that $\Delta r_{t-3}(s)$ terms are excluded. However, restrictions of this sort are only valid if the VAR in terms of the two interest rates has exactly the order specified, but because it is only an approximation, we estimate unrestricted VAR's for r_t^* . We illustrate the cross-equation restrictions implied by the expectations theory for a second-order system. The extension to higher-order VAR's follows directly.

2. The Markov-switching approach used by Sola and Driffill (1994) has the advantage over the threshold approach that closed-form solutions for multistep predictions are easily obtained, essentially because the Markov-chain state variable is exogenous. However, the exogeneity of the regime-switching process is at the same time a drawback if we wish to model the dependence of the dynamic relationship between the spread and the short rate on the level of the spread directly.

3. The computational burden of estimating the three-regime TVAR is reduced by adopting a "one-step-at-a-time" approach, following Hansen (2000), whereby the estimated threshold value of

a two-regime model is taken as one of the thresholds of the three-regime model, and a grid search for the second threshold is then conducted. This procedure can be iterated to improve the estimate of the first threshold, and so on. We adopt this procedure for both nonlinear models that we estimate. Bai (1997) proved the consistency of this approach for models with multiple structural breaks.

4. Of course, if policy is credible, the long rate should reflect the expected future low-inflation environment brought about by higher short rates. A time-varying risk premium could be another possible explanation for the negative correlation between the spread and the long-term interest rate. Hardouvelis (1994) finds little support for this, while Tzavalis and Wickens (1998) take the opposite position.

REFERENCES

- Anderson, H.M. (1997) Transaction costs and non-linear adjustment towards equilibrium in the US Treasury bill market. *Oxford Bulletin of Economics and Statistics* 59, 465–484.
- Bai, J. (1997) Estimating multiple breaks one at a time. *Econometric Theory* 13, 315–352.
- Balke, N.S. & T.B. Fomby (1997) Threshold cointegration. *International Economic Review* 38, 627–645.
- Campbell, J.Y. (1995) Some lessons from the yield curve. *Journal of Economic Perspectives* 9, 129–152.
- Campbell, J.Y. & R.J. Shiller (1987) Cointegration and tests of present value models. *Journal of Political Economy* 95, 1062–1088.
- Clements, M.P. & A.B. Galvão (2001) A Comparison of Tests of Nonlinear Cointegration with an Application to the Predictability of US Interest Rates Using the Term Structure. Mimeo, University of Warwick.
- Clements, M.P. & J. Smith (1997) The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting* 13, 463–475.
- Dwyer, G.P., P. Locke, & W. Yu (1996) Index arbitrage and nonlinear dynamics between the SP500 futures and cash. *Review of Financial Studies* 9, 301–332.
- Enders, W. & C.W.J. Granger (1998) Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business and Economics Statistics* 16, 304–311.
- Ericsson, N.R. & J. Marquez (1998) A framework for economic forecasting. *Econometrics Journal* 1, C228–C266.
- Granger, C.W.J. & T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.
- Gray, S.F. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27–62.
- Hall, A.D., H.M. Anderson, & C.W.J. Granger (1992) A cointegration analysis of Treasury bill yields. *Review of Economics and Statistics* 74, 116–126.
- Hamilton, J.D. (1988) Rational-expectations econometric analysis of changes in regime. An investigation of the term structure of interest rates. *Journal of Economics, Dynamics and Control* 12, 385–423.
- Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
- Hansen, B.E. (2000) Testing for linearity. In D.A.R. George, L. Oxley, & S. Potter (eds.), *Surveys in Economic Dynamics*, pp. 47–72. Oxford: Blackwell.
- Hansen, B.E. & B. Seo (2001) (in press). Testing for two-regime threshold cointegration in vector error correction models. *Journal of Econometrics*.
- Hardouvelis, G.A. (1994) The term structure spread and future changes in long and short rates in the G7 countries. *Journal of Monetary Economics* 33, 255–283.
- Mankiw, N.G. & J.A. Miron (1986) The changing behaviour of the term structure of interest rates. *Quarterly Journal of Economics* 101, 211–228.
- Martens, M., Kofman, P., & T.C.F. Vorst (1998) A threshold error-correction model for intraday futures and index returns. *Journal of Applied Econometrics* 13, 245–263.

- Pagan, A.R., A.D. Hall, & V. Martin (1996) Modelling the term structure. In G.S. Maddala & C.R. Rao (eds.), *Handbook of Statistics, Statistical Methods in Finance* Vol. 14: pp. 91–118. Amsterdam: North-Holland.
- Pfann, G.A., P.C. Schotman, & R. Tschernig (1996) Nonlinear interest rate dynamics and implications for the term structure. *Journal of Econometrics* 74, 149–176.
- Roberds, W., D. Runkle, & C.H. Whiteman (1996) A daily view of yield spreads and short-term interest rate movements. *Journal of Money, Credit and Banking* 28, 34–53.
- Rudebusch, G.D. (1995) Federal reserve interest rate targetting, rational expectations, and the term structure. *Journal of Monetary Economics* 35, 245–174.
- Sola, M. & J. Driffill (1994) Testing the term structure of interest rates using a stationary vector autoregression with regime switching. *Journal of Economic Dynamics and Control* 18, 601–628.
- Tsay, R.S. (1998) Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93, 1188–1202.
- Tzavalis, E. & M. Wickens (1998) A re-examination of the rational expectations hypothesis of the term structure: Reconciling the evidence from long-run and short-run tests. *International Journal of Finance and Economics* 3, 229–239.
- van Dijk, D. (1999) *Smooth Transition Models: Extensions and Outlier Robust Inference*. Ph.D. Dissertation, Tinbergen Institute.
- van Dijk, D. & P.H. Franses (2000) Nonlinear error-correction models for interest rates in the Netherlands. In W.A. Barnett, D.F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim, & A. Wurtz (eds.), *Nonlinear Econometric Modelling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory*. Cambridge, UK: Cambridge University Press.
- Watson, M.W. (1999) Explaining the increased variability in long-term interest rates. *Federal Reserve Bank of Richmond, Economic Quarterly* 85, 71–96.

APPENDIX A

To evaluate the RHS of (10), note that

$$E_{t-1}\Delta r_{t+j}(1) = i_r E_{t-1}\mathbf{R}_{t+j} = i_r \left[\mathbf{C} \sum_{s=0}^j \mathbf{\Gamma}^s + \mathbf{\Gamma}^{j+1} \mathbf{R}_{t-1} \right]$$

so that the RHS of (10) can be written as (omitting θ):

$$\frac{1}{k} \left[\sum_{i=1}^{k-1} \sum_{j=1}^i E_{t-1}\Delta r_{t+j}(1) \right] = \frac{1}{k} i_r \sum_{j=1}^{k-1} (k-j) \left[\mathbf{C} \sum_{s=0}^j \mathbf{\Gamma}^s + \mathbf{\Gamma}^{j+1} \mathbf{R}_{t-1} \right].$$

If we focus solely on the term in \mathbf{R}_{t-1} ,

$$\frac{1}{k} i_r \sum_{j=1}^{k-1} (k-j) \mathbf{\Gamma}^{j+1} \mathbf{R}_{t-1} = i_r \mathbf{\Gamma}^2 (\mathbf{I} - \mathbf{\Gamma})^{-1} (\mathbf{I} - \mathbf{\Gamma}^{k-1}) \mathbf{R}_{t-1} - \frac{1}{k} i_r \mathbf{\Gamma} \sum_{j=1}^{k-1} j \mathbf{\Gamma}^j \mathbf{R}_{t-1}.$$

For large k , the term in \mathbf{R}_{t-1} is approximately $i_r \mathbf{\Gamma}^2 (\mathbf{I} - \mathbf{\Gamma})^{-1} \mathbf{R}_{t-1}$, because $\mathbf{\Gamma}^{k-1} \simeq 0$, and $j \mathbf{\Gamma}^{j+1} \rightarrow 0$ [see, e.g., Ericsson and Marquez (1998, p. C239)].

APPENDIX B

Writing the elements of Γ_1 and Γ_2 defined in (7) as

$$\Gamma_1 = \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix}, \quad \Gamma_2 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}$$

(when $p = 2$), the cross-equation restrictions implied by equation (13) for the VAR of the spread and the short rate when $k = 2$ are

$$\varphi_{11}(2 - \varphi_{21}) = \varphi_{22}\varphi_{21} + \theta_{21}$$

$$\varphi_{12}(2 - \varphi_{21}) = \varphi_{22}\varphi_{22} + \theta_{22}$$

$$\theta_{11}(2 - \varphi_{21}) = \varphi_{22}\theta_{21}$$

$$\theta_{12}(2 - \varphi_{21}) = \varphi_{22}\theta_{22}.$$

When the stronger version of the theory is employed, the restrictions are

$$\varphi_{21} = 2, \quad \varphi_{22} = 0, \quad \theta_{21} = 0, \quad \theta_{22} = 0$$

For $k = \infty$, the restrictions given by equation (11) are

$$\varphi_{11}(1 - \varphi_{11} - \varphi_{21}) = \varphi_{22}\varphi_{21} + \varphi_{12}\varphi_{21} + \theta_{11} + \theta_{21},$$

$$\varphi_{12}(1 - \varphi_{11} - \varphi_{21}) = \varphi_{22}\varphi_{22} + \varphi_{12}\varphi_{22} + \theta_{22} + \theta_{12},$$

$$\theta_{11}(1 - \varphi_{11} - \varphi_{21}) = \varphi_{22}\theta_{21} + \varphi_{12}\theta_{21},$$

$$\theta_{12}(1 - \varphi_{11} - \varphi_{21}) = \varphi_{22}\theta_{22} + \varphi_{12}\theta_{22},$$

and for the stronger version,

$$1 - \varphi_{11} = \varphi_{21}$$

$$-\varphi_{12} = \varphi_{22}$$

$$\theta_{11} + \theta_{21} = 0$$

$$\theta_{12} + \theta_{22} = 0.$$